City of Urbana, Illinois – Traffic Stop Data Analysis

2007 through 2009

Tom Christoff

The following report performed by Tom Christoff includes analyses of the Illinois Department of Transportation (IDOT) data for the City of Urbana for the years 2007, 2008 and 2009. This report was prepared at the request of Todd Rent in his capacity as Human Relations Officer for the City of Urbana. The data used in the following analysis were collected as part of the Illinois Traffic Stops Statistics Study[1]. As requested by Mr. Rent, the following report aims to analyze the following relationships:

- The relationship between race and the number of stops.
- The relationship between beat assignment and the population of stopped drivers with specific regards to the stopped driver's race, age and gender.
- The relationship between the number of calls for service in a specified beat and the number of stops in that beat.
- The relationship between the number of crime reports in a specified beat and the number of stops in that beat.
- The relationship between Beat 61 and the above factors.
- The relationship between time of day and the number of stops.
- The relationship between day of the week and the number of stops
- The relationship between the stopped driver's age and the number of stops
- The relationship between the stopped driver's gender and the number of stops.
- The relationship between age of vehicle and the rate of ticketing (particularly for equipment violations).

The purpose of this report is to view what role race plays in the decision to stop a motorist as well as what role race plays in the decision to give a warning or citation to the stopped motorist. Along with race, the other factors above were analyzed in order to see whether they play a similar or larger role than race in these decisions.

The following report describes statistical analyses which some explanation of is necessary for the lay person. Attached as Appendix J of this report is a Definition Guide which provides definitions for all processes which may be referred to if there is any confusion with any of the terms used in this report. A quick overview will be provided here, explaining terms and processes used for each analysis.

Each analysis has an Independent and Dependent variable. An independent variable is one that does not change as a result of another variable. A dependent variable is one that does change as a result of another variable. For example, if fire touches my finger, I move my finger. Therefore, my finger is dependent on the fire as to whether I will move it. The fire does not depend on my finger as to whether it will burn something. My finger (or more accurately, my movement of the finger) is the dependent variable, while the fire is the independent variable.

[1] The Illinois Department of Transportation Traffic Stop Reports are yearly published reports detailing city, town, municipality and county information regarding the number of traffic stops made upon various racial/ethnic groups. These stops are compared with the racial composition of the area in order to view whether the percentage of minority stops in an area are in proportion to the racial composition. The City of Urbana, for the 2009 report, had a minority stop ratio of 1.56, meaning that for every 1 Caucasian, 1.56 minorities were stopped. In the 2010 report, the City of Urbana had a minority stop ratio of 1.7

Most of the variables used in the tests were recoded for analytical reasons. To recode a variable simply means to give it a different name or provide it a numerical assignment. For example, for the variable Result of Stop, the possible result of the stop (warning given/citation given) were recoded and provided numerical values – Warnings = 0, while Citation = 1. This was done to reflect the notion that Citations were a larger penalty and therefore given a larger number.

Each analysis begins with an examination of the variables' frequencies. A frequency is a statistical output detailing the number of times a specific variable entry is found and the percentage of times this number appears compared with other variable entries. If you count to 10, each number is counted only once (its numerical frequency is 1) and represents 1/10 (10%) of the numbers. If the number four was included twice, then the numerical frequency for the number four would be 2 and it would represent 2/11 (18%) of the data set.

The variables are also examined to determine their correlation with the result of the stop. Correlation means whether one variable has a relationship with another. It should be strongly noted that no inferences should be made by this analysis. One well known rule in statistics is that correlation does not equal causation. For example, let's say we were looking at shoe size and wondering whether it correlates with math comprehension. It would not be surprising to find that smaller shoe sizes do not have the level of math comprehension as bigger shoe sizes, as older individuals would have a larger shoe size and probably a better comprehension of mathematics. Rather, correlation shows a relationship without considering any other factors (such as age in the math example). Still, it provides a starting point to looking at the relationship of variables.

There are two types of possible relationships which variables can have with each other: a positive relationship or a negative relationship. The term "positive relationship" means that as one variable increases, the other variable also increases. The term "negative relationship" means that as one variable increases, the other variable decreases. Think of it as multiplying two numbers. If one is increasing (+) and the other is decreasing (-), the result is negative. If one is increasing (+) and the other is also increasing (+), the result is positive. If one is decreasing (-) and the other is also decreasing (-), the result is still positive.

The strength of the relationship is determined by its statistical significance. The term "statistical significance" means "a number that expresses the probability that the result of a given experiment or study could have occurred purely by chance". In statistical analysis, the highest acceptable level of significance is .10. Therefore, any number over .10 is considered insignificant, or unacceptably chance related.

The above terms and processes are ones which are found in each analysis undertaken. One analysis (The Relationship Between Age of Vehicle/Equipment Violation Stops and the Rate of Ticketing) contains further statistical tests which are explained in that section. However, for the majority of analyses, the above explanations should be sufficient to understand the discussion of the results.

The Relationship Between Race and Stops

When looking at the relationship between Race and whether an individual was stopped, the results should first be looked at in its totality. As seen in Table 1, Whites were the most stopped race in the entire data set, compromising 53.8% of all stopped motorists. As was expected, Blacks made up the second highest group with 33.8 percent of all stopped motorists. Hispanics made up 4% of the total stops and Asians were stopped 8.3% of the time, while American Indians made up zero percent of the stops (meaning stops made upon this race made up less than .05% of all stops).

Table 1 – Race recoded for all five categories

	Frequency	Percent	Valid Percent	Cumulative Percent
White	6291	53.8	53.8	53.8
Black	3950	33.8	33.8	87.6
Hisp	472	4.0	4.0	91.7
Asian	969	8.3	8.3	100.0
AI/AN	4	.0	.0	100.0
Total	11686	100.0	100.0	

It is also interesting to look at the race of an individual and the difference in whether they will receive a citation or a warning. In a perfect distribution, we would expect to find that the amount of citations as compared to warnings would be equal between each race. That is to say that if one race receives 40% citations and 60% warnings, this would also be true for every other race. Looking at Appendix A, this does appear to be true for the majority of races.

From Appendix A, we see that Whites, Blacks and Asians all show an approximate 40% citation rate while receiving warnings approximately 60% of the time. While American Indians show a large difference from this (75% citation and 25% warning), this should not be examined too closely due to the fact that they only had a total of 4 total stops, which is less than .005% of all stops effected. Hispanic show a larger percentage of citations as opposed to warnings and the possible cause for this is discussed below. However for all other races, the percentages of citations and warnings are consistent in their application.

Using a secondary data set provided by the City of Urbana detailing the number of citations which were given for offenses that have a higher likelihood of resulting in tickets, we see that Hispanics have a higher percentage of stops in this category than any other race. The data set includes three types of offenses which are more likely to receive citations than other categories: No Driver's License, Breath Alcohol Over Limit, and Operating Uninsured Vehicle. These three offenses will more likely to result in a citation than less serious offenses (i.e. failure to signal, speeding less than five miles per hour over the speed limit, etc). While this data set and the one used for all other analyses in this report should not be

considered the same, they are comparable in that they encompass the same time period, the same police agency and the same racial groups.

When looking at the number of occurrences in which Hispanics are cited for the above reasons, we see that Hispanics have a total 278 citations for these reasons. Taking the total of 278 and dividing it into the total number of citations Hispanics received in the original data set (357), the results indicate that approximately 77.87% of all Hispanic citations are the result of one of the above listed offenses (278/357 = 77.87%). When looking at all other races and their percentage of citations which come from the above reasons, a much smaller number is found (1711/6550 = 26.12%). Therefore, while there may appear to be a racially charged difference in the percentage of citations given to Hispanics, this is most probably due to the fact that Hispanics are more likely to be cited for the above reasons as compared to all other races.

The Relationship Between Police Beat and Stops

Before running analysis on the beats, the data had to be converted from geo-codes (five digit indicators of where the stop occurred) into beats (two digit indicators of a larger area). After conversion, it was found that a total of 259 were considered missing because they were associated with geo-codes that did not correspond with any of the beat numbers. These cases constituted approximately 2.2 percent of the entire data set. I am satisfied that the loss of this 2.2 percent will not affect the analysis in any significant way.

The analysis of each beat was undertaken in five separate ways:

1. The relationship of each beat with the race of the individual stopped
2. The relationship of each beat with the gender of the individual stopped
3. The relationship of each beat with the age of the individual stopped.
4. The relationship of each beat with the number of stops compared to the number of Calls for Service
5. The relationship of each beat with the number of stops compared to the number of Crime Reports

Before looking at these associations, a view of the frequency of stops for each beat with allow for a preliminary picture to be seen. Looking at Table 2, it can be seen that the highest percentage of stops occurred in Beat 61 (26.6 percent of all stops). Beat 65 had the lowest percentage of stops (16.4 percent of all stops), while Beats 62 – 64 were in the middle range. One note regarding this Table is the difference between “Percent” and “Valid Percent”. This difference is due to the loss of the aforementioned 259 cases. The “Valid Percent” column does not take these missing cases into consideration when formulating the percentages. The “Percent” column does take these missing cases into account. For the purposes of this report, the column “Valid Percent” will be used for analysis.

Table 2 - Beat Recoded into Larger Beat Districts

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	61	3039	26.0	26.6	26.6
	62	2078	17.8	18.2	44.8
	63	2291	19.6	20.0	64.8
	64	2148	18.4	18.8	83.6
	65	1871	16.0	16.4	100.0
	Total	11427	97.8	100.0	
Missing	System	259	2.2		
Total		11686	100.0		

Turning now to the five relationships stated above, we may begin with the relationship of each beat to the race of the stopped individual (See Attached Microsoft Excel Worksheet). There are two things to note while examining the Excel Worksheet. First, it is unknown based on the data what the racial residential composition of each Beat is. The data only indicates who was stopped in the Beat, not who lives there. For the purposes of this analysis, it is assumed that racial composition is distributed proportionately in each Beat as to the composition in the entire city. It is up to you the reader to interpret the results in the manner you see fit in light of your understanding of the racial proportionality of each Beat. Secondly, when looking at the "Percentage Within Race (or Gender or Age)" column in the Excel Worksheet, it must be remembered that Beat 61 had a higher percentage of stops than all other Beats and therefore will have a higher percentage of stops within each group. For example, should all Beats had the same amount of stops and all races were stopped proportionately, we would expect the "Percentage Within Race" column to be consistently 20%, since there are five Beats. So, an even distribution would show that 20 percent of White stops were in Beat 61, 20 percent of White stops were in Beat 62, etc. Since Beat 61 had a higher frequency of stops, then a higher percent of White stops would be in that Beat. The opposite holds true for Beat 65, which had a lower frequency of stops.

When looking at the Excel Worksheet, we do find that Beat 61 has a generally higher percentage of stops within each race than other Beats and that Beat 65 has a generally lower percentage of stops within each race than other Beats. As explained above, this is to be expected. However, in Beats 62 – 64, we do not find the proportionality which would be expected if we were to assume even residential distribution. For example, in Beat 63, it is shown that 35.6 percent of all Asian/Pacific Islander stops occurred in this beat, while approximately 1/10 of all Black stops occurred in this beat. Again, this must be interpreted with the reader's knowledge of the residential racial composition of this beat.

What may be stated is the beat where each race is stopped the most and with your knowledge, you may interpret whether this is consistent with the racial makeup of each beat. White stops throughout the data were fairly evenly distributed with the most stops coming in Beat 63 (23.1%), while Beats 64 and 65 show a slightly less percentage of stops (16.3% and 18.3%, respectively). With respect to Black stops,

more than 1/3 of all Black stops occurred in Beat 61, with Beat 64 holding 24.7% of all Black stops. Beat 63 held the lowest amount of Black stops with 11.8%. Looking at Hispanic stops, a fairly even distribution can be seen with Beats 63 and 65 showing lower percentages (14.8% and 14.3%, respectively). Beats 61, 62 and 64 show Hispanic stop percentages in the lower to middle 20's. In terms of Asian/Pacific Islander stops, we find a large proportion (35.6%) in Beat 63, while Beat 61 holds 28.6% of their stops. The other three beats are each responsible for approximately 12% of Asian/Pacific Islander stops. The Native American/Alaskan Native group only had four stops in the entire data set. Only Beat 63 did not have a stop of this group, while all other beats had 1.

Unlike race, we would not expect gender to differ significantly from beat to beat. While it is not uncommon for cities to have areas with higher percentages of minority or White populations, it would be uncommon for a city to have an area that was mostly male or mostly female. Therefore, we would expect a fairly even distribution of male and female stops within each beat. The results of the gender analysis show this to be true. As expected, due to the frequency of stops in Beats 61 and 65, we find that Beat 61 has the highest amount of male/female stops and Beat 65 has the lowest amount of male/female stops. The other three beats show a percentage around 20 percent for each gender. In all beats, males were stopped more times than females; however the proportion of males to females remains fairly consistent.

Due to the presence of a University, we may expect that the distribution of stops as a result of age will differ from beat to beat. Some neighborhoods may be more inclined to house University students, while others are more likely to house permanent city residents. For example, almost 50% of all stops in Beat 61 came from Groups 14 and 15 (Age 20-29). Looking at the Within Age column, approximately 1/3 of all stops in Group 15 (Age 25-29) were stopped in Beat 61. On the other hand, Groups 2-4 (Age 75-89) only had approximately 10 percent of their total stops occur in this beat. However, in Beat 65, approximately half of the stops for Group 2 (Age 85-89) occurred here.

Looking at the results beat by beat, each beat essentially follows the same pattern of increasing in the number of stops as the Group number increases (as the stopped individual becomes younger) before decreasing when reaching Group 17 (Age 15-19). Should we expect an even distribution for all age groups across all beats, we would expect to see that each age group has 20% of their stops in each beat, since there are 5 beats. We do not see this, indicating that there is not equal distribution across all beats.

For Beat 61, it appears that most younger people are stopped here, with Groups 8-16 (Age 15-59) and specifically Groups 13-16 (Age 15-34), showing near or more than one quarter of their stops occurring in this beat. As stated before, approximately 10 percent of stops of more elderly individuals occur in this beat. For Beat 62, we find closer to the 20% percent distribution of stops we would expect if all stops were equally distributed. Group 3 (Age 80-84) is underrepresented in this beat with only 13.6 percent of stops is this age group occurring in this beat. Groups 15 and 16 also lose some percentage in this beat compared with Beat 61. Of interest is that 36.4 percent of Group 1 (Age 90-95) occurs in Beat 62, though with only 11 total stops, this amount should not be scrutinized too closely. For Beat 63, we again find numbers near the 20% even distribution mark. Individuals in Group 3 (Age 80-84) were more

likely to be stopped in this beat (31.4% of Group 3 stops occurred in this beat), though they only had 51 total stops. Overall though, we find even distribution for groups in this beat. For Beat 64, it appears that for individuals in Groups 2 and 3 (Age 80-89), only approximately 5% of their stops occurred in this beat. For Group 1 (Age 90-95), we again see a high percentage of their stops occurring in this beat, though for the small amount of total stops, this again should not be scrutinized too closely. The results show smaller percentages for Groups 5, 6, and 8 (Age 65-74 and 55-59, respectively). All other groups show relatively even distribution. For Beat 65, we find that older people who are stopped are more likely to be stopped in this beat. This beat had the least amount of stops in the entire data set, and although more younger people are likely to be stopped here (the same as with all other beats), proportionately they are less likely to be stopped in this beat. It may probably be assumed that this beat encompasses an area that is least likely to be associated with student living.

Looking at the beats individually, the question may be asked if there is a difference between each beat and the result of the stop (i.e. whether a warning or citation is issued). Looking at the crosstabulation between the beats and result (Appendix B), the results indicate that overall, approximately 60% of stops receive a citation, while approximately 40% of stops receive a warning. However, there is some variation in this amount between beats. For example, Beat 62 has a fairly even distribution between warnings and citations (49.5% and 50.5%, respectively). The strictest beats are Beats 61 and 65, which give citations approximately 64% of the time. Beats 63 and 64 show that citations are given approximately 56% of the time, while warnings are given approximately 43% of the time.

The relationship between a specific beat's Calls for Service and Crime Reports and their amount of stops is also important to look at since there may be more police patrolling an area with more crime or more calls for service. Information provided by the City of Urbana shows the amount of Calls for Service and Crime Reports in each beat location for the years 2007 through 2009. This amount of Calls and Crime for each beat is listed below for all three years. The first analysis done is with the combination of all three years to determine whether there is a relationship between the amount of stops made in each beat and the amount of Calls for Service/Crime Reports for each beat. After that, each year will be examined individually. We will start with the amount of Calls for Service and then look at Crime Reports.

CALLS FOR SERVICE

	2007	2008	2009	Total
Beat 61	4285	3606	3448	11339
Beat 62	3601	3642	3800	11043
Beat 63	4368	4599	4433	13400
Beat 64	4708	4361	4293	13362
Beat 65	3644	4129	4026	11799

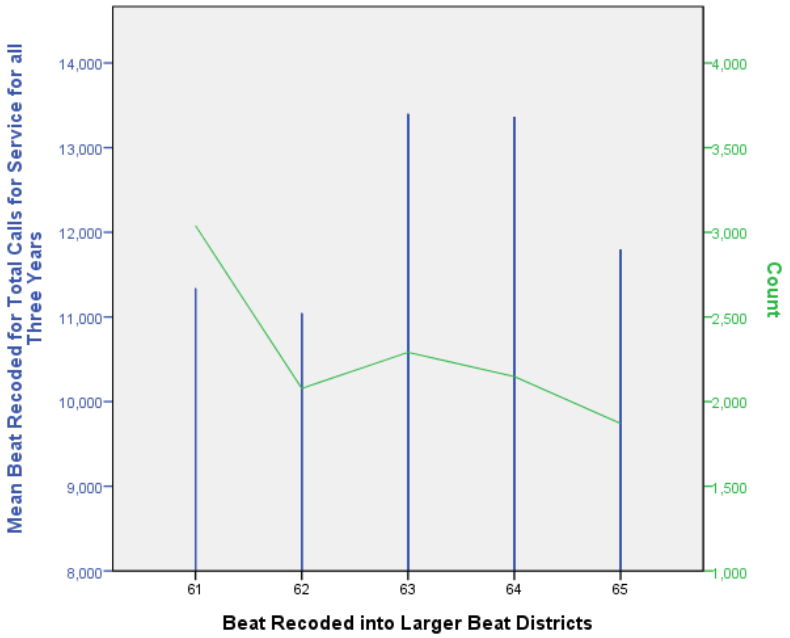
After determining the total amount of Calls for Service, these numbers were input for each time the beat occurred, first for all three years combined, then for each individual year. For example, each time Beat 61 occurred in the data, a separate variable that was created showed the number 11339 (the total number of Calls for Service over all three years). For the individual years, the number of Calls for Service for that year was inputted. For example, each time Beat 61 occurred in the data *and* the year of stop was 2007, a separate variable that was created showed the number 4285 (the number of Calls for Service in 2007). With this, it could be seen if the number of Calls for Service was proportionate to the amount of stops made by each beat.

We would expect to see that a beat with a high number of Calls for Service would also have a high number of stops, and a low number of Calls for Service would have a low number of stops. The results indicate that this is to some extent true. Looking at the graph for all three years (Graph 1), we see a decent amount of correlation with the number of stops and the Calls for Service. The blue lines in the graph indicate the amount of Calls for Service for each beat, while the green line indicates the amount of stops made in each beat. By far the highest amount of stops was made in Beat 61, where the second lowest amount of Calls for Service was made. However, should we disregard Beat 61, we see that the pattern of stops follows the pattern of Calls for Service for the rest of the beats.

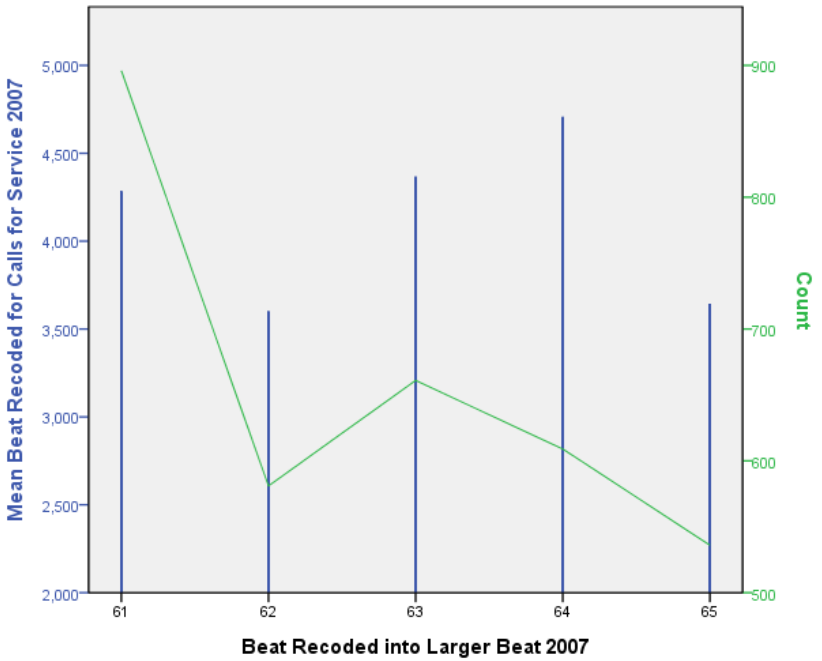
This result also holds true when looking at the year by year analysis (Graphs 2 through 4). Again, Beat 61 , in all three years has, the most number of stops, but in none of those three years does that beat have the most Calls for Service. Should we take away Beat 61 though, the pattern of stops follows the pattern of Calls for Service. In 2007, Beat 64 does not conform to the pattern, but other than that instance and Beat 61, the pattern fits. Also of interest is the difference between Calls for Service in 2007 for Beat 61 and Calls for Service in Beat 61 in the other two years. There is an approximately 600 Calls

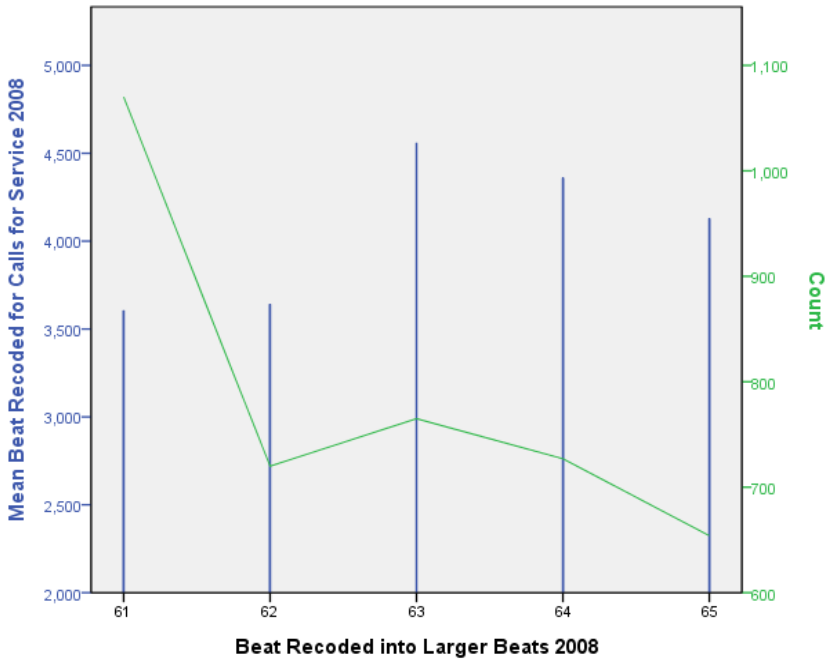
for Service decrease for Beat 61 between 2007 and 2008/2009, though there may not be any specific reason for this.

Graph 1

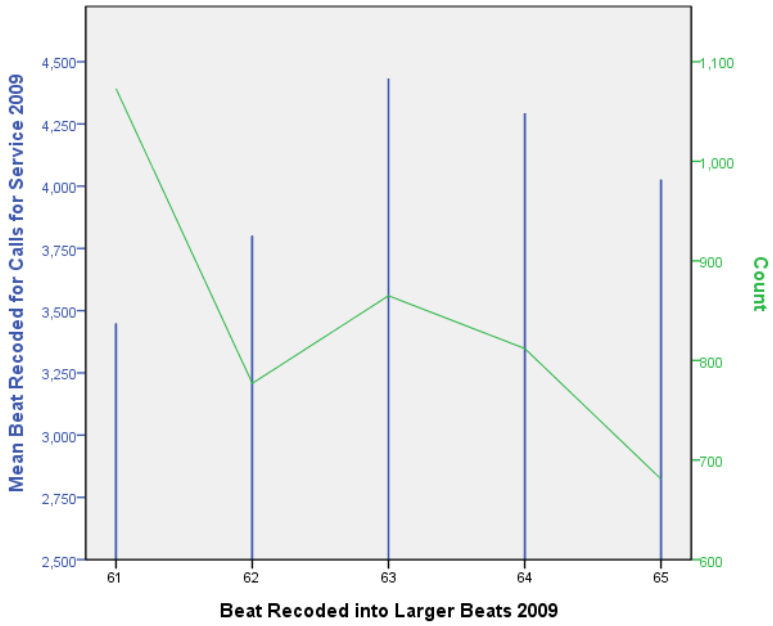


We may also look at whether there is a relationship between Calls for Service and the result of the traffic stop. To do this, the data was recoded to reflect the order the beats fell into in terms of the amount of Calls for Service. For example, for all three years, Beat 62 had the least amount of Calls for Service and was therefore assigned the number "1", while Beat 63 had the highest amount of Calls for Service and was therefore assigned the number "5". With this, we can see if an increase in Calls for Service led to more or less citations or warnings given. Looking at the crosstabulations (Appendix C), we don't necessarily find much relationship with the amount of Calls for Service and the result of the stop. For instance, the beat with the lowest amount of Calls for Service had the most warnings compared to the amount of citations it gave out, but the beat with the highest amount of Calls for Service had the second most warnings compared to the amount of citations it gave out. The second highest amount of Calls for Service had the third highest percentage of warnings given compared to citations. Therefore there is not much relationship between these two variables.





Graphs 2 though 4



CRIME REPORTS

	2007	2008	2009	Total
61	1458	1265	1135	3858
62	1276	1238	1180	3694
63	1306	1259	1192	3757
64	1843	1669	1701	5213
65	1264	1461	1263	3988

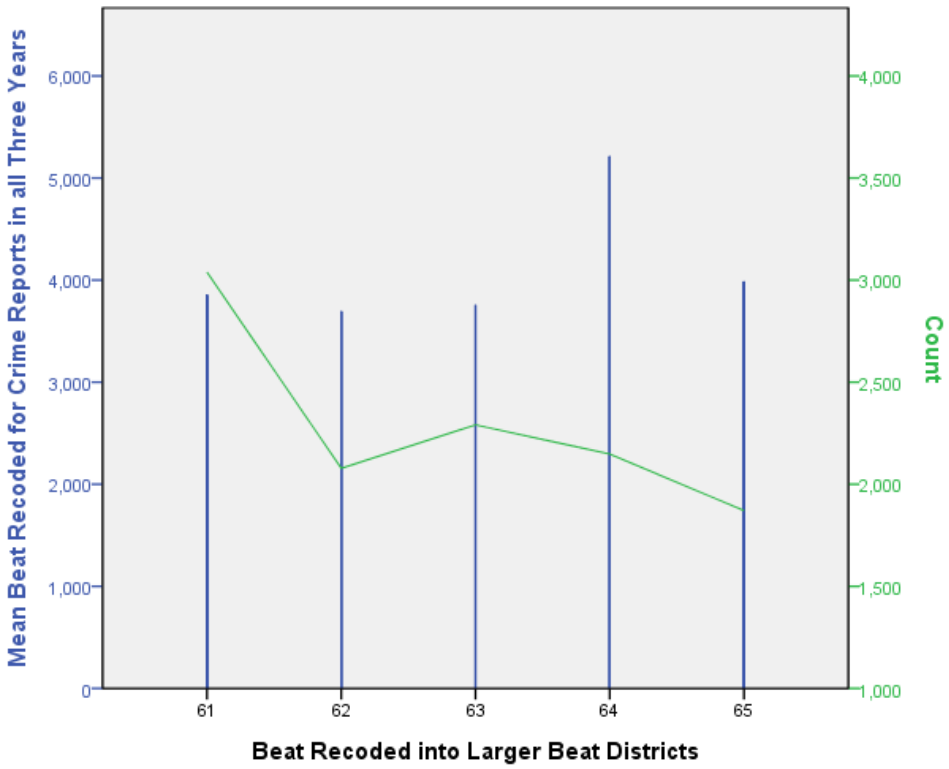
Looking now at the Crime Reports and the relationship with the amount of stops made for all three years, we find less correlation between the numbers than found for Calls for Service (Graph 5). Again, though Beat 61 had fewer Crime Reports, it had by far the most stops. Also, the number of Crime Reports for Beat 64 does not match its proportion with the number of stops in that Beat. It may be said that the relationship between Crime Reports and the number of stop within each beat is not as strong as with Calls for Service, at least for all three years combined.

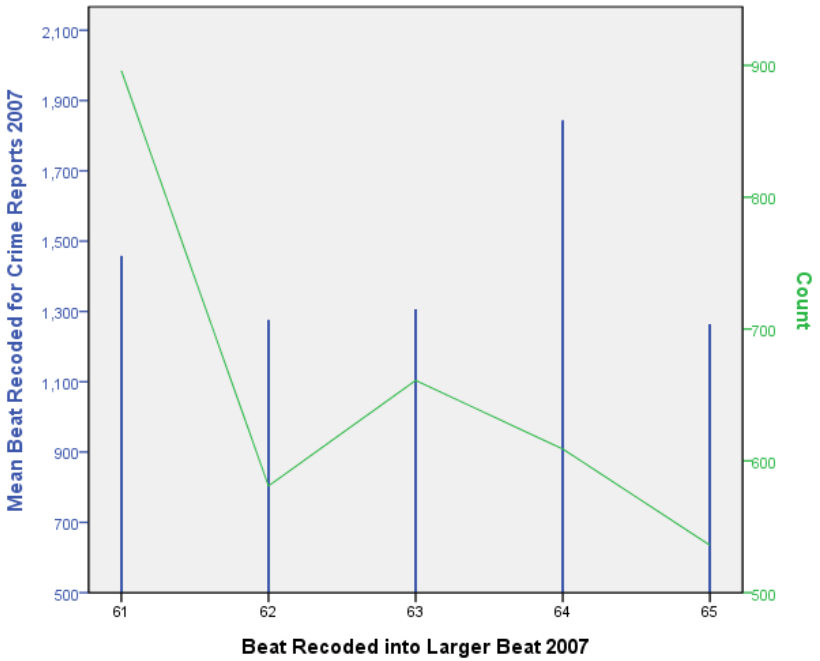
Looking at the graphs for all three years (Graphs 6 through 8), we consistently find Beat 64 having by far the highest amount of Crime Reports, but also consistently find that this is not reflected in the amount of stops which occur in that beat. As with Calls for Service, Beat 61 has a disproportionately high amount of stops compared to the number of Crime Reports there. As shown with the graph detailing all three years combined, the graphs showing the year by year comparison of Crime Reports and the number of stops shows less of a relationship than the analysis of Calls for Service and the number of stops in a beat.

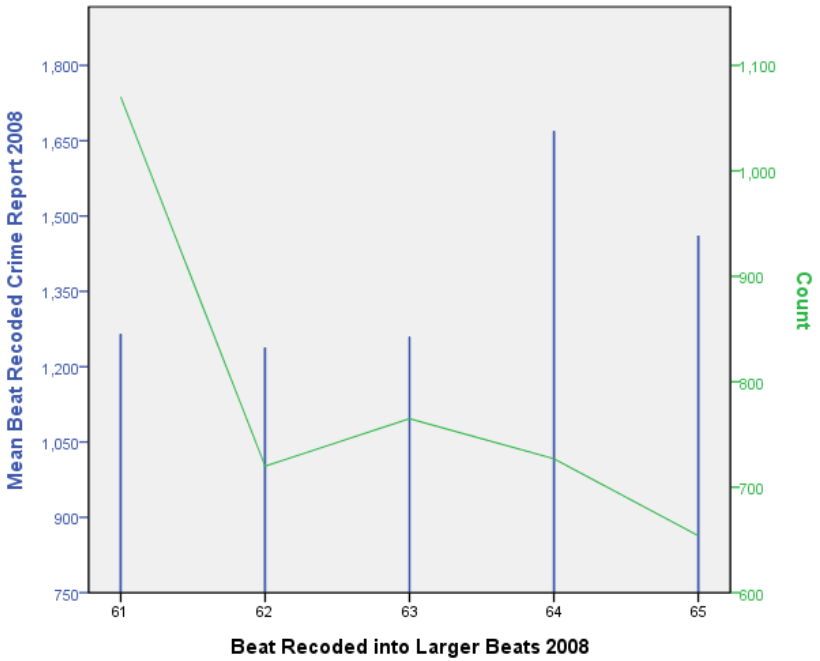
When examining the relationship between the amount of Crime Reports and the result of the ticket (Appendix D), the beats were again assigned a number based on their order for the number of Crime Reports, with the beat with the highest number of Crime Reports being assigned the number “5” and the beat with the lowest number of Crime Reports being assigned the number “1”. Here, we find a small relationship as the amount of citations increases from positions 1-3. However, positions 4 and 5 see a decrease in the amount of citations written compared to the amount of warnings. So while there may

be a relationship in the first three positions, this relationship reverses when it comes to the last two positions.

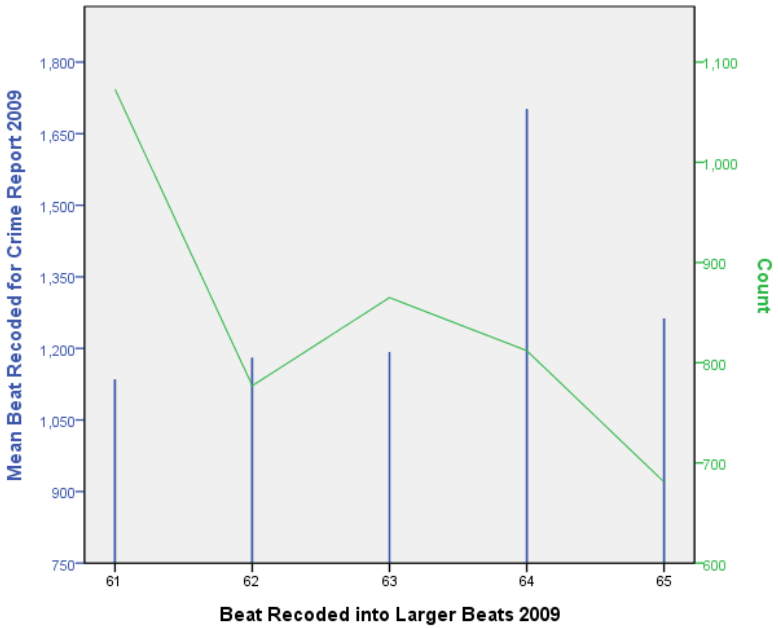
Graph 5







Graphs 6 though 8



Examination of Beat 61 Individually

The data can also be broken down to exclude Beat 61 to understand whether the data as a whole is more reflective of the general population if this Beat is not included. When looking at the complete data set (including Beat 61), the following percentages are found regarding the race of individuals who were stopped:

- White: 54.1%
- Black: 33.5%
- Hispanic: 4.0%
- Asian: 8.3%
- American Indian: 0% (Denotes that the percentage is less than .005%)

When removing Beat 61 from analysis, we are also removing approximately 26.6% of the entire data set from the analysis. When this Beat is removed, the data is changed and the following percentages are found regarding individuals who were stopped:

- White: 57.4%
- Black: 30.3%
- Hispanic: 4.1%
- Asian: 8.1%
- American Indian: 0%

Looking at the individual numbers for *only* Beat 61, the following percentages are found regarding individuals who are stopped:

- White: 45.0%
- Black: 42.3%
- Hispanic: 3.7%
- Asian: 9.0%
- American Indian: 0%

Based on these numbers it is seen that there are differences between the race of individuals stopped in Beat 61 and the rest of the larger beats in general; however these differences are more concerned with Whites and Blacks in their variations. The difference between Hispanics, Asians, and American Indians are small and are probably due to chance. Yet the differences between the stops of Whites and Blacks

are larger and should be looked at closer. Beat 61 shows stops for both races almost equally (45% for Whites and 42.3% for Blacks), while the difference in the other Beats shows 57.4% for Whites and 30.3% for Blacks. The other Beats more accurately reflect the data as a whole (54.1% for Whites and 33.5% for Blacks).

Also, looking at the ratio for minority stops compared to the estimated minority driving population, there is a difference in what these numbers would have been, had Beat 61 been excluded from the analysis. In 2007, the ratio was 1.47; in 2008, the ratio was 1.49; in 2009, the ratio was 1.56. Adding these three ratios together and dividing by three (giving us the average ratio) results in an average ratio of 1.51. If Beat 61 is excluded from this analysis, the ratio drops to 1.39 for all three years. It can be assumed then that Beat 61 has a higher percentage of minority stops and is raising the total ratio which is being reported to the Illinois Department of Transportation. However, even with the exclusion of Beat 61, the percentage of minority stops is above their percentage within the general population, though this may be due to the fact that the estimated minority driving population is based on census data from the year 2000.

To further investigate the reason that Beat 61 has a differing percentage of minority (particularly Black) stops than all the rest of the Beats, Beat 61 was broken down into the Geo-Codes which were originally part of the data set provided. Overall, there were 20 individual Geo-Codes within Beat 61, though a number of these Geo-Codes had less than 25 total stops within all three years. For the purposes of this report, only Geo-Codes which had twenty-five or more total stops within all three years are included in this analysis. This led to a total of 13 Geo-Codes being eligible for analysis. The attached Microsoft Excel Spreadsheet shows the difference between stops based on race, gender and age within each Geo-Code in Beat 61. Since the data for each of the 13 Geo-Codes are included in the Excel Spreadsheet, we may only focus on Geo-Codes with very high stop amounts within the text of this report.

The Geo-Code with the highest amount of stops, 54606, had a total of 580 stops over the three year course. The racial breakdown for stops in this Geo-Code shows 47.9% of the stops were made on Whites, while 33.8% of the stops were made on Blacks and 15% on Asians. Only 3.3% of the stops were made on Hispanics. This is fairly consistent with the racial breakdown of stops made in the data set as a whole (see below). The gender variation is also consistent with the data as a whole (see below). Also, the age distribution follows a similar pattern as the data on the whole (see below) with the number of stops increasing as the age group gets younger.

The Geo-Code with the second highest amount of stops was 51808 with a total of 463 stops for the three year span. This Geo-Code had similar results to 54606 when looking at gender and age distributions, though the distribution of race was more even in terms of Whites and Blacks. White's composed 47.7% of this Geo-Code's stops while Blacks composed 42.5% of the stops. This is also different from the distribution of race in the complete data set.

Geo-Codes 54607 and 54610 have approximately the same number of stops (383 and 364, respectively) in the data set and both Geo-Codes show similar distributions. Both Geo-Codes show more than 50% of their stops were made on Whites, though Geo-Code 54610 has a larger percentage of Black stops

(35.2%). Geo-Code 54607 has a more even distribution of stops with regards to gender, though men are still stopped at a higher rate. As with other Geo-Codes (and with the data as a whole) both 54607 and 54610 show an increase in stops as the age group of the stopped individual becomes younger.

The above described Geo-Codes have a large number of stops compared to other Geo-Codes in Beat 61. For all other Geo-Codes, see the attached Excel Spreadsheet. Most of the Geo-Codes follow similar patterns, however large differences from the standard pattern should be examined with consideration of the number of stops made in the area as well as the racial composition of the Geo-Code.

The Relationship Between Time of Day and Stops

Before looking at the relationship between time of day and the number of stops, the data was recoded to reflect the number of seconds which the time represented. Similar to military time, this recoding allows the data to be viewed in a linear fashion rather than two sets of 12 hour blocks (AM and PM). For example, the time 12:00:00 AM, would be represented by the number "0", as it would be the beginning of the day. The time 11:59:59 PM, which would be the last possible time before the new day (and therefore a new count) would be represented by the number "43199". This is calculated by breaking up the time into hours, minutes and seconds. The hours and minutes must be converted into the amount of seconds they represent. The following calculations show how this is done:

11 (hrs) X 60 (number of minutes in each hour) X 60 (number of seconds in each minute) = 39600

59 (mins) X 60 (number of seconds in each minute) = 3540

59 seconds = 59 seconds

39600 + 3540 + 59 = 43199

After performing these calculations on all times provided in the data set, the numbers were grouped together to form six (6) time groups. These time groups were labeled 1-6 with the number 1 representing early morning and the number 6 representing late night. The groups are as follows:

4:00 AM – 7:59 AM – 1

8:00 AM – 11:59 AM – 2

12:00 PM – 3:59 PM – 3

4:00 PM – 7:59 PM – 4

8:00 PM – 11:59 PM – 5

12:00 AM – 3:59 AM – 6*

*While this is technically considered morning, the reporting investigator believed it better serving to include this range as a later number. As the progression goes from morning to night, it is the belief of the reporting investigator that 2:00 AM, for example, would still be considered "night" for the individual being stopped.

A frequency of the stops based on these groups of time is included first in this report (Table 3). This frequency is found below.

Table 3 - Time in Seconds Recoded Into Six Groups

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	567	4.9	4.9	4.9
	2	1595	13.6	13.6	18.5
	3	1922	16.4	16.4	35.0
	4	1921	16.4	16.4	51.4
	5	3904	33.4	33.4	84.8
	6	1776	15.2	15.2	100.0
	Total	11685	100.0	100.0	
Missing	System	1	.0		
Total		11686	100.0		

When looking at the results of the frequency output (Table 3), we can clearly see two groups which are extremely different from the others. By far, the lowest number of traffic stops occurred in Group 1 (4:00 AM – 7:59 AM). Only 567 traffic stops occurred during this time frame over the three years analyzed, constituting only 4.9 percent of all traffic stops in the data. On the other extreme, the highest number of traffic stops occurred in Group 5 (8:00 PM – 11:59 AM). This time frame constituted approximately 1/3 of all traffic stops in the data. All other time groups are approximately the same in terms of the number of stops which occurred within them.

Also of interest is the combination of Group 5 and Group 6. When looking at the Cumulative Percent column, we can see that 51.4 percent of all traffic stops were made between Groups 1 and 4 (between 4:00 AM and 7:59 PM). This means that 48.6 percent, or approximately half, of all the traffic stops in the data were made between 8:00 PM and 3:59 AM. Of even higher percentage is the combination of Group 4 and Group 5 (4:00 PM to 11:59 PM). These groups account for 49.8 percent of all the traffic stops in the data.

We may also look at whether the time of day bears any significance upon the result of the encounter. In order to do this, the data was recoded to account for whether the individual received a warning (either written or verbal) or whether they received a citation. In this analysis, citations are represented by the

number “2”, while written and verbal warnings are represented by the number “1”. This coding was done to represent the notion that citations are larger penalties and as such were assigned a larger number. When looking at the results (Appendix E), we find that from 4:00 AM to 8:00 PM, citations are written at least 2/3 of the time, with Groups 1 and 3 approaching approximately 75 percent of the time. Interestingly, in Groups 5 and 6 (8:00 PM to 3:59 AM), this trend reverses, with more warnings than citations given. Although the difference in result is small (approximately 52 percent warnings to 48 percent citations), the difference from Groups 1 through 4 is apparent.

The Relationship Between Day of the Week and Stops

The data for this analysis was drawn from the Date of Stop variable in the data set. Using Microsoft Excel, the day of the week which correlated to the date of the stop was calculated. The day was then recoded in order to reflect the order of the week which the day was situated. For example, Monday was given the number 1, Tuesday was given the number 2, etc., ending on Sunday which was given the number 7. The data was also recoded for weekdays and weekends. The days Friday, Saturday, and Sunday were given the number 1 while all other days were given the number 0. This allows for an analysis of each day as well as the more general group of days known as weekends and weekdays.

Looking at the frequencies (Table 4) and graph (Graph 9), we find a fairly consistent number of stops occurring throughout the week, with each day (except Sunday) being responsible for approximately 15 percent of the stops. Monday shows a slightly lower percentage (13.4 percent), though this is not a significant difference. Sunday by far shows the least amount of stops, reflecting only 9.7 percent of all stops in the data. This is also reflected when looking at the data coded for weekdays and weekends only, with the weekdays showing approximately 60 percent of all stops in the data and the weekends holding 40 percent.

We may also look at whether the day of the week has any relationship with the outcome of the stop. Interestingly, the results indicate that there is a relationship. For the days Monday through Friday (Appendix F), the results show that approximately 62 percent of all stops are given citations, while approximately 38 percent are given some type of warning. Saturday and Sunday are quite different though. On Saturdays, the percentage of tickets written drops to 53.6 percent, while warnings increase to 46.4 percent. When looking at Sunday, the trend switches, with more warnings given than tickets (52.5 percent compared to 47.5 percent, respectively). This is also reflected in the correlation (Table 5), as well as the correlation for the weekend/weekday recoded variable (Table 6) showing a significant negative relationship (that is to say, as the weekend occurs, more warnings than tickets are given).

Table 4 - Day of Week Recoded for the Location in Week

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	M	1564	13.4	13.4	13.4
	Tu	1812	15.5	15.5	28.9
	W	1814	15.5	15.5	44.4
	Th	1810	15.5	15.5	59.9
	F	1746	14.9	14.9	74.8
	Sa	1802	15.4	15.4	90.3
	Su	1138	9.7	9.7	100.0
	Total	11686	100.0	100.0	

Graph 9

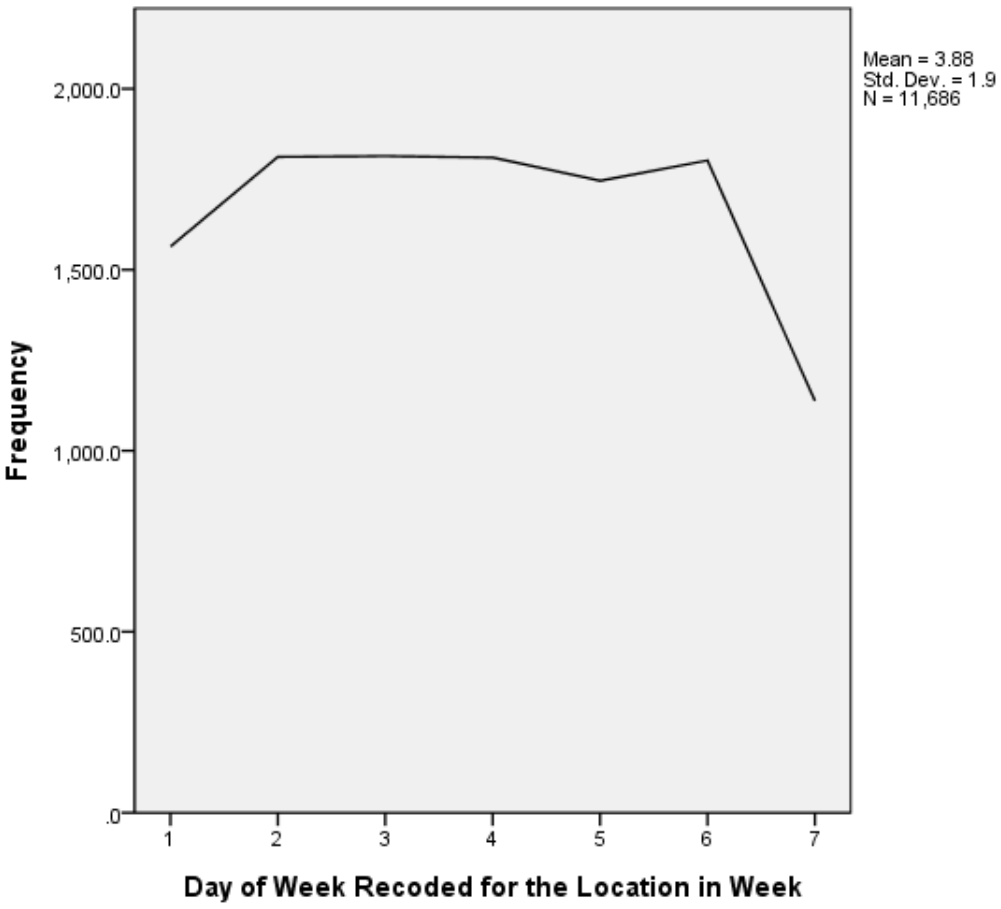


Table 5 – Correlations

		Result of Stop Recoded	Day of Week Recoded for the Location in Week
Result of Stop Recoded	Pearson Correlation	1	-.078**
	Sig. (2-tailed)		.000
	N	11686	11686
Day of Week Recoded for the Location in Week	Pearson Correlation	-.078**	1
	Sig. (2-tailed)	.000	
	N	11686	11686

**. Correlation is significant at the 0.01 level (2-tailed).

Table 6 – Correlations

		Day of Week Recoded for Weekend	Result of Stop Recoded
Day of Week Recoded for Weekend	Pearson Correlation	1	-.068**
	Sig. (2-tailed)		.000
	N	11686	11686
Result of Stop Recoded	Pearson Correlation	-.068**	1
	Sig. (2-tailed)	.000	
	N	11686	11686

**. Correlation is significant at the 0.01 level (2-tailed).

The Relationship Between Driver Age and Stops

Before looking at the relationship between the number of stops and the age of the driver, the data was recoded into sixteen groups of ages. Rather than analyze the data for each individual driver (a span of 80 years), driver's ages were grouped together in groups of five (5) year intervals. For example, the earliest year of birth in the data set was 1913. This individual was stopped in 2008. Subtracting the year 1913 from when the individual was stopped, the result shows the individual was stopped when he or she was 95 years old. Thus the age group 90-95 was created. While this is a six year age span, all other groups are five year intervals. So, the next group of ages is 86-89. Individuals in the eldest category are represented by the number "1" in the data set. Individuals aged 15-19 (the youngest age group) are represented by the number "16" in the data set. A key showing the ages of the individuals in each group is provided below.

Looking at the frequency output (Table 7) and the graph (Graph 10) below, it can be said quite convincingly that the age of an individual is related to being stopped. As the age of those stopped goes down (that is, as they get younger) the number of stops continuously increases. In fact, there is not one instance in the data where a younger group is stopped less than an older group (except for Group 16, Age 15-19 years old).

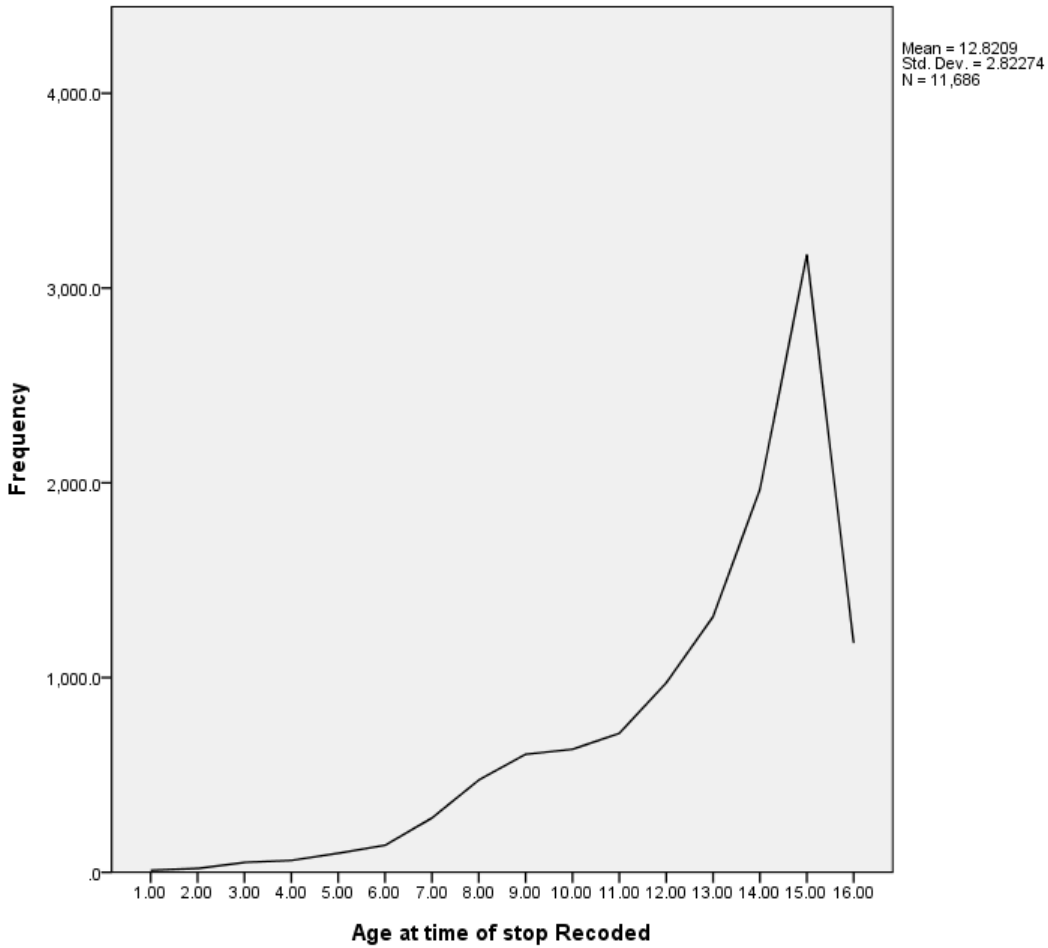
Indeed, the highest percentage of stops was made on individuals who were between the ages of 20 and 24. This group of individuals constituted approximately one quarter of all individuals stopped. When combining Groups 13, 14 and 15 (individuals stopped between the ages of 20 and 34), approximately 55 percent of all stops were made on individuals in these groups. This is not surprising considering the presence of a major university in the immediate area. Also, it has been shown in past research that younger drivers are more likely to engage in risky driving behaviors and are therefore more likely to be stopped as a result of this.

It was also desired to know whether a stopped individuals age lead to a higher risk for receiving a citations versus receiving a warning. Looking at Appendix G, it is quite obvious that younger drivers are more likely to receive a citation rather than a warning. When looking at Groups 1 and 2 (Age 85-95), the results show that citations are written approximately 35 percent of the time, while warnings are given the remaining approximate 65 percent of the time. Granted these are a small percentage of the total stops; however, after Group 3 (Age 80-84), which had an inexplicably large jump to 52.9 percent citations written, the groups follow a fairly specific pattern. Groups 4, 5 and 6 (Age 65-79) received citations between 40-49 percent of the time, while receiving warnings the rest of the time. Groups 7-12 (Age 35-64) received citations between 50-59 percent of the time. Finally, Groups 13-16 received citations between 60-63 percent of the time. The correlation table (Table 8) confirms this, showing a significant positive relationship between Age and Result of Stop. As the Age Group number increases (as the stopped motorist becomes younger) the Result of Stop also increases (more citations than warnings).

Table 7 - Age at time of stop Recoded

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	11	.1	.1	.1
	2.00	20	.2	.2	.3
	3.00	51	.4	.4	.7
	4.00	61	.5	.5	1.2
	5.00	98	.8	.8	2.1
	6.00	139	1.2	1.2	3.3
	7.00	279	2.4	2.4	5.6
	8.00	474	4.1	4.1	9.7
	9.00	606	5.2	5.2	14.9
	10.00	632	5.4	5.4	20.3
	11.00	714	6.1	6.1	26.4
	12.00	973	8.3	8.3	34.7
	13.00	1313	11.2	11.2	46.0
	14.00	1967	16.8	16.8	62.8
	15.00	3171	27.1	27.1	89.9
	16.00	1177	10.1	10.1	100.0
	Total	11686	100.0	100.0	

Graph 10



1 = 90-95 4 = 75-79 7 = 60-64 10 = 45-49 13 = 30-34 16 = 15-19

2 = 85-89 5 = 70-74 8 = 55-59 11 = 40-44 14 = 25-29

3 = 80-84 6 = 65-69 9 = 50-54 12 = 35-39 15 = 20-24

Table 8 – Correlations

		Age at time of stop Recoded	Result of Stop Recoded
Age at time of stop Recoded	Pearson Correlation	1	.079**
	Sig. (2-tailed)		.000
	N	11686	11686
Result of Stop Recoded	Pearson Correlation	.079**	1
	Sig. (2-tailed)	.000	
	N	11686	11686

**. Correlation is significant at the 0.01 level (2-tailed).

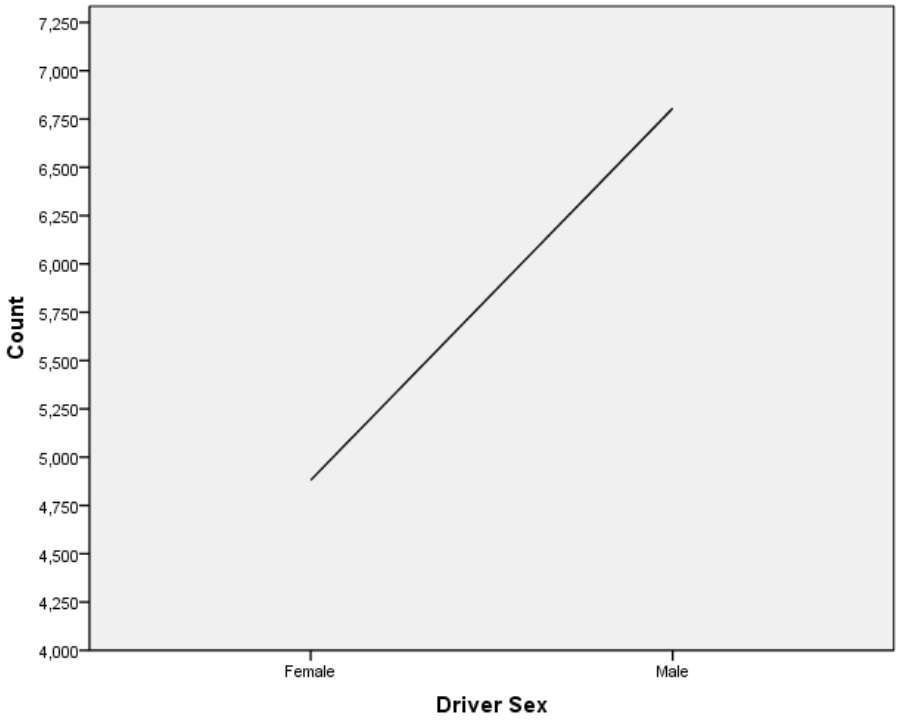
The Relationship Between Gender and Stops

Finally, we can examine the relationship between a stopped motorists gender and being stopped. For this analysis, the variable gender was recoded in order to view the correlation between gender and the result of the stop. Therefore, males are represented by the number "1" while females are represented by the number "2". Before examining the correlations, the frequency for the gender variable and the graph showing the frequency are provided. The frequency (Table 9) and graph (Graph 11) show that males make up approximately 60% of all stops, while females are responsible for approximately 40% of all stops.

Table 9 - Driver Sex

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Female	4881	41.8	41.8	41.8
	Male	6805	58.2	58.2	100.0
	Total	11686	100.0	100.0	

Graph 11



We should also look at the correlation between the gender of the stopped motorist and whether they received a citation or warning. Consistent with the findings in the first section, it is seen in the correlation table (Table 10) that there is an insignificant relationship between the driver's gender and whether they receive a citation or not. So while males are stopped at a higher percentage than women, there does not appear to be any difference in whether one gender will receive a citation compared to the other gender.

Table 10 – Correlations

		Sex Recoded	Result of Stop Recoded
Sex Recoded	Pearson Correlation	1	.011
	Sig. (2-tailed)		.236
	N	11686	11686
Result of Stop Recoded	Pearson Correlation	.011	1
	Sig. (2-tailed)	.236	
	N	11686	11686

The Relationship Between Age of Vehicle/Equipment Violation Stops and the Rate of Ticketing

Before looking at the relationship between the age of the vehicle being stopped/those vehicles being stopped for equipment violations and the rate of ticketing, the data was recoded for both independent variables (Age of Vehicle and Equipment Violations) and the dependent variable (Result of Stop). The variable Age of Vehicle was recoded for the sake of simplicity. Rather than run analysis on every single year a car was made in, the data was recoded for groups of five year intervals. This yielded nine groups. Vehicles made in 1971 or earlier are represented by the number "1". Vehicles made between the years 1972 – 1976 are represented by the number "2". The groups continue in this fashion, with the last group of vehicles, those made in 2007 or later, being represented by the number "9".

Upon review of the data, only three reasons for stops were provided: Equipment Violations, Moving Violations, and License Plate/Registration. The data was then recoded to account for only equipment violations. Equipment violations are therefore represented by the number "1", while all other reasons for stops are represented by the number "0".

While the concern of the relationship in question dealt with the result of the stop, it is still worthwhile to examine the frequencies regarding the independent variables and being stopped in of itself. When looking at the frequency for Age of Vehicle (Table 11), the vast majority of vehicles stopped in the data set were made between the years 1992-2010 (Groups 6-9). These years constituted approximately 90 percent of all vehicles which were stopped. The highest percentage for a single group in the Year of Vehicle variable was Group 7. Approximately 1/3 of all stops were made on vehicles in Group 7, which are vehicles made between the years 1997-2001.

Turning now to the frequencies for Reason for Stop (Table 12), they are listed as they were before they were recoded. Based on this output, we can see that individuals are stopped for Equipment Violations approximately 25% of the time. Moving Violations and License Plate/Registration Violations make up the remaining 75% of the stops, with Moving Violations make up approximately 2/3 of the total stops (67.6%).

Finally, the frequency for Result of Stop (Table 13) indicates that on a whole, citations are given to stopped motorists approximately 59% of the time while receiving a warning the rest of the time.

Table 11 - Year of Vehicle Recoded for Every Five Years

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1.00	5	.0	.0	.0
	2.00	16	.1	.1	.2
	3.00	56	.5	.5	.7
	4.00	182	1.6	1.6	2.2
	5.00	834	7.1	7.1	9.4
	6.00	2460	21.1	21.1	30.4
	7.00	3875	33.2	33.2	63.6
	8.00	3204	27.4	27.4	91.0
	9.00	1053	9.0	9.0	100.0
	Total	11685	100.0	100.0	
Missing	System	1	.0		
Total		11686	100.0		

1 = >1971 4 = 1982 – 1986 7 = 1997 - 2001

2 = 1972 – 1976 5 = 1987 – 1991 8 = 2002 – 2006

3 = 1977 – 1981 6 = 1992 – 1996 9 = 2007 – 2010

Table 12 - Reason for Stop

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Equipment	2928	25.1	25.1	25.1
	License Plate/Regist	862	7.4	7.4	32.4
	Moving Violation	7896	67.6	67.6	100.0
	Total	11686	100.0	100.0	

Table 13 - Result of Stop

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Citation	6907	59.1	59.1	59.1
	Written Warning	4779	40.9	40.9	100.0
	Total	11686	100.0	100.0	

The correlations for the independent variables (Table 14) are broken down into two groups. The independent variable Year of Vehicle and its relationship to the Result of Stop on the Correlations chart below is designated by "**". The independent variable Reason for Stop and its relationship to the Result of Stop on the Correlations chart below is designated by "##".

Table 14 – Correlations

		Result of Stop Recoded	Year of Vehicle Recoded for Every Five Years	Reasons for stop recoded for equipment
Pearson Correlation	Result of Stop Recoded	1.000	.008**	-.259##
	Year of Vehicle Recoded for Every Five Years	.008**	1.000	-.146^^
	Reasons for stop recoded for equipment	-.259##	-.146^^	1.000
Sig. (1-tailed)	Result of Stop Recoded	.	.188**	.000##
	Year of Vehicle Recoded for Every Five Years	.188**	.	.000^^
	Reasons for stop recoded for equipment	.000##	.000^^	.
N	Result of Stop Recoded	11685	11685	11685
	Year of Vehicle Recoded for Every Five Years	11685	11685	11685
	Reasons for stop recoded for equipment	11685	11685	11685

With regards to the Year of Vehicle (designated by **), the output of the Pearson Correlation shows that there is a positive relationship between the two variables (.008 is a positive number with a value greater than zero). The statistical significance of this correlation (Sig. (1-tailed)) is .188. The value of .188 in this model is then not significant.

With regards to the Reason for Stop (designated by ##), the output of the Pearson Correlation shows that there is a negative relationship between the two variables (-.259 is a negative number with a value less than zero). The statistical significance of this correlation (Sig. (1-tailed)) is .000. This means that the strength of the correlation is significantly not likely to be an affect of chance.

For illustrative purposes, look at the correlation between Year of Vehicle and Reason for Stop (designated by ^^). The output of the Pearson Correlation shows a negative relationship (-.146). This is saying as a vehicle gets newer (the value *increases* from 1 to 9), the reason for the stop is less likely to be for an equipment violation (the value *decreases* from 1 to 0). This is not surprising as newer vehicles

should not have as many equipment violations as older vehicles. We would expect there to be a negative relationship then.

With our knowledge of positive and negative relationships, an examination of regression is possible. The term regression refers to the effect that one variable (the independent variable) has on another variable (the dependent variable). This is different from a correlation in that a correlation merely looks at the relationship between the two – regression looks at the influence that the independent variable has on the dependent variable. As with correlation, nothing concrete can be determined from one simple analysis. There should be no expectation of any causal effect. It takes many analyses to come away with an explanation of cause. Think of the relationship between smoking cigarettes and lung cancer. No one analysis can prove smoking cigarettes causes cancer. Rather, you must take the aggregate of all studies done over an extended course of time to state that causal relationship. Therefore, since this report only relies on one data set and one analysis, nothing should be taken for granted.

The tables below show the regression analysis for the independent variables Year of Vehicle and Reason for Stop on the dependent variable Result of Stop. When looking at the individual independent variables in a regression analysis, it must be understood that the other independent variable is being controlled for. For example, the output for Reason for Stop is shown with the variable Year of Vehicle controlled for. This will be explained in the next paragraph.

When looking at the Reason for Stop (independent variable) regression on the Result of Stop (dependent variable) (Table 15), we can see there is a negative relationship (t=-3.363) (designated with ***). As explained above, this negative relationship means that as the Reason for Stop increases (more stops for equipment violations), the Result of Stop decreases (more warnings than citations). This is with the Year of Vehicle being controlled for. Controlling for Year of Vehicle means that the analysis is taking away any possibility that the Year of Vehicle will influence the Result of Stop – that is to say that no matter how old the vehicle is, if a person is stopped for an equipment violation, the result of the stop will more likely be a warning than a citation.

When looking at the Year of Vehicle (independent variable) regression on the Result of Stop (dependent variable) (Table 15), we now see a negative relationship (t=-29.198) (designated with ###). That is to say as the Year of Vehicle increases (as the vehicle gets newer), the Result of Stop decreases (more warnings than citations). This switch between positive and negative relationships is due to the fact that we are now controlling for the Reason for Stop and because we are looking at effect rather than simple correlation. So no matter whether a person is stopped for an equipment violation or another reason (the control of the variable), a newer vehicle will more likely receive a warning than a citation.

For both of these independent variables, the statistical significance is within the acceptable frame (remember that statistical significance is the probability that the results are due to chance). These values are designated by ^^^ in Table 15. As you can see, both values are under .10 (the highest acceptable level). The strictest statistical significance value marker used in statistical analysis is .01,

which both of these values are under. We can therefore be confident that these results are not due to chance.

The Model Summary (Table 16) is important to examine due to the Adjusted R Square results. Simply stated, the Adjusted R Square states how much of the variation of the dependent variable (Result of Stop) can be explained by the independent variables (Year of Vehicle and Reason for Stop). Remember the example from before about the relationship between shoe size and math comprehension. Remember how there would be a positive relationship which would be significant. While the example was a correlation example, it would also hold true for a regression. However, the Adjusted R Square would be low in the example, as the amount of variation in math comprehension explained by shoe size would be small. Put frankly, some other variable would explain more of the variation (such as age or quality of education). This will make more sense in the Step-Wise analysis provided later. In the current analysis, the Adjusted R Square is .068, meaning that only 6.8 percent of the variation in the result of the stop can be explained by these two independent variables. So while we can say that these variables significantly explain whether someone will receive a citation or a warning (the variance), it only explains about 6.8 percent of the variance. Other variables are at work in explaining the variation.

Table 15 - Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	1.755	.027		64.416	.000	1.701	1.808
	Year of Vehicle Recoded for Every Five Years	-.013	.004	-.030	-3.363***	.001^^^	-.020	-.005
	Reasons for stop recoded for equipment	-.299	.010	-.264	-29.198###	.000^^^	-.319	-.279

a. Dependent Variable: Result of Stop Recoded

Table 16 - Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.261[a]	.068	.068	.475

a. Predictors: (Constant), Reasons for stop recoded for equipment, Year of Vehicle Recoded for Every Five Years

The following Step-Wise analysis allows the variables to be compounded on top of each other to view how the dependent variable is affected. In the Step-Wise, there are multiple models. The first model shows one independent variable's affect on the dependent variable. The second model adds an independent variable and shows those two variables affect on the dependent variable. The third model adds another independent variable, and so forth.

For the analysis of Year of Vehicle and Reason for Stop, we have already seen that the inclusion of these two variables does not adequately explain the variance in whether a ticket or warning is given. That is to say that neither of these two variables, when including and controlling for the other, adequately explains the outcome of the stop. Should we include other variables, more of the variation can be explained (See Appendix H).

When looking at the Step-Wise (Appendix H), we can jump to Model 3, as Model 1 only includes Year of Vehicle (which results in an insignificant relationship) and Model 2 is the same as the analysis explained above (Year of Vehicle and Reason for Stop as an Equipment Violation as the Independent Variables). Model 3 includes Race in the analysis. For the purposes of analysis, the variable Race was recoded so that Caucasians are represented by the number "0" while minorities are represented by the number "1".

When looking at the coefficients of the Step-Wise output for Model 3, the variable Race has positive relationship (t=7.823). That is to say as a person becomes "more minority" (since minority is 1 and Caucasian is 0), they will receive a higher penalty (i.e. a citation as opposed to a warning). The relationship is also statistically significant (sig. = .000). The analysis of the variable Race in this model is also holding the variables Year of Vehicle and Reason for Stop constant. That is to say, no matter how old or new a vehicle or whether the stop was due to an equipment violation, minorities are significantly more likely to receive a ticket rather than a warning.

Yet when looking at the Adjusted R Square of Model 3 (Table 17), it is shown that these three variables only explain 7.3 percent of the variation in Result of Stop. So while we may say that minorities are more likely to receive tickets than warnings when controlling for Year of Vehicle and Reason for Stop, there is something else that has not been accounted for that will explain the variation in whether the stopping officer will issue a ticket or a warning.

Model 4 includes whether the Reason for Stop was a Moving Violation (moving violation is represented by the number "1" while all other reasons are represented by the number "0"). Even with the inclusion of this variable, the model (Table 17) still only explains 8.2 percent of the variation in the Result of Stop. Again, there is something else influencing the result of the stop. Normally in statistical analysis, an acceptable Adjusted R Square value lies somewhere between 20 and 50 percent. Obviously there are going to be factors that cannot be accounted for (for example, the car had a White Sox bumper sticker and the stopping officer is a Cubs fan). However, if you can explain 20 to 50 percent of the variation in social science statistics, you can claim to explain a significant portion of the variation. Since Model 4 (the model with the most variables represented) only explains 8.2 percent of the variation, it can be determined that some other factor (or factors) is influencing a police officer's decision to issue a ticket or warning.

Table 17 - Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.008[a]	.000	.000	.492
2	.261[b]	.068	.068	.475
3	.270[c]	.073	.073	.473
4	.286[d]	.082	.082	.471

a. Predictors: (Constant), Year of Vehicle Recoded for Every Five Years

b. Predictors: (Constant), Year of Vehicle Recoded for Every Five Years, Reasons for stop recoded for equipment

c. Predictors: (Constant), Year of Vehicle Recoded for Every Five Years, Reasons for stop recoded for equipment, Race Recoded

d. Predictors: (Constant), Year of Vehicle Recoded for Every Five Years, Reasons for stop recoded for equipment, Race Recoded, Reason for Stop Recoded for Moving Violation

We can also examine how all the variables examined in this report affect an officer's decision to issue a warning or citation. Looking at Appendix I, all variables which are examined in this report are included as independent variables while the Result of Stop is included as the dependent variable. The only variable which does not reach significance at any level is the Sex of the stopped motorist. Therefore, any variation in whether a male or a female receives a warning or citation is unacceptably possibly due to chance. All other variables are significant at either the .05 level or the .01 level, indicating that they are all significantly related to whether an individual receives a warning or citation. However, looking at the Model Summary (Appendix I), the Adjusted R Square is only .108, indicating that only 10.8 percent of the variation in whether a person receives a warning or citation can be explained by the inclusion and control of all these variables. Again, there must be other factors which must be included to adequately explain the variance in the Result of Stop.

Finally, returning to the relationship between Age of Vehicle and whether the individual was stopped for an Equipment Violation (as Independent Variables), and the Result of Stop (as the Dependent Variable), the following Binary Logistic computation was created to view whether the Result of Stop was affected if an individual was stopped in an older vehicle AND for an equipment violation. For this analysis, the variable Year of Vehicle was recoded to become a binary variable (two options) – vehicles older twenty years old are represented by the number "1", while vehicles that are newer than twenty years old are represented by the number "0". A Binary Logistic test compounds the variables, examining only cases in which both variables are in represented by the number "1". Remember that a vehicle over twenty years old is represented by the number "1" and equipment violations are represented by the number "1". Therefore, only cases that have a "1" for both variables are included in the analysis.

The results are interesting in that we have already seen that an older vehicle is less likely to receive a citation compared to a warning and that equipment violations are less likely to receive a citation as well. When looking at the results (Table 18), we see that recoding of the vehicle into older or newer than twenty years turns the relationship insignificant for this individual variable, though still a negative relationship. The relationship between a vehicle being stopped for an equipment violation is still

significantly negatively related to whether they receive a citation. However, when combining these two variables (a vehicle is over twenty years old AND is stopped for an equipment violation), the result is significantly positively related to receiving a citation. This relationship is significant at the .05 level. Looking at the Model Summary (Table 19), the Cox and Snell R Square is similar to the Adjusted R Square and indicates that only approximately 6.5% of the variation can be explained by this model.

Table 18 - Variables in the Equation

		B	S.E.	Wald	Df	Sig.	Exp(B)
Step 1[a]	ReasonRECEquip	-1.271	.048	709.370	1	.000	.281
	YearREC.Twenty.Years	-.023	.083	.078	1	.780	.977
	Both	.415	.137	9.130	1	.003	1.514
	Constant	.686	.024	845.991	1	.000	1.987

a. Variable(s) entered on step 1: ReasonRECEquip, YearREC.Twenty.Years, Both.

Table 19 - Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	15019.007[a]	.065	.088

a. Estimation terminated at iteration number 3 because parameter estimates changed by less than .001.